YOONBO CHO

AI Engineer · Applied LLM Systems

Elkins Park, PA · +1 718 541 6008 · yoonbocho1@gmail.com · yoonbo1.github.io · github.com/yoonbo1 · linkedin.com/in/yoonbo-cho-9780301b9 · Remote US, Philadelphia or NYC; relocation to Chicago or WA

SUMMARY

AI engineer, 4+ years. Former Senior Developer at A View From My Seat (**1.2M monthly active users**, native iOS and Android, 80+ releases). Since 2026, building LLM products solo: Everblue (live on the App Store) and tickatap.com.

WORK EXPERIENCE

Senior Developer — A View From My Seat — Nov 2024 – Jul 2026 · Philadelphia, PA

*Sports/venue platform · **1.2M monthly active users***

- Owned native iOS (Swift) and Android (Kotlin) development end-to-end: spec, architecture, implementation, QA, and release, shipping **80+ updates** at feature parity across platforms.
- Shipped **vision-based moderation** for user-submitted seat-view photos, auto-flagging or clearing each image for publication, and backfilled structured metadata like view obstructions across the photo corpus.
- Built a **venue-scouting agent** that runs daily, diffs live venue data against stored snapshots, and alerts only on real changes, replacing manual monitoring.

Founder & AI Engineer · Everblue — Yoonibo LLC — Jan 2026 – Present

*AI journaling iOS app · **everblue.me** · live on the App Store, launched Jun 2026*

- Built an **LLM evaluation harness** against the production pipeline using synthetic day payloads. Used a manual release gate of **92%+ generation quality and 100% no-fabrication** on the evaluated cases.
- Built **on-device hybrid retrieval** with MiniLM ONNX embeddings, BM25, recency decay and MMR reranking to retrieve months of journal context without a vector database.
- Built a **Django LLM backend** with feature-based provider routing, byte-identical prompt fallback, retries, PII redaction and per-device request budgets.
- Shipped **Florence-2 Core ML photo captioning** through a Swift bridge and resolved memory crashes on 4GB iPhones. Later moved captioning to a gated backend service, removing **~1.1GB from the app bundle**.

Founder & AI Engineer · tickatap — Yoonibo LLC — Jul 2026 – Present

*Invitation-first going-out planner · **tickatap.com***

- Shipped **tickatap.com**, an invitation-first going-out planner, solo. Built a TypeScript platform with **Next.js, Fastify, Expo and Postgres**; friends choose real events and places to assemble a shared plan.
- Built a daily **event-curation agent** combining local-model extraction with frontier-model curation. Added source validation, cron locks and freshness checks to prevent stale or invented events.
- Enforced booking authorization in the **server-side tool executor** and verified it with a 25-case adversarial suite. Retired the booking agent after it achieved **13.3% task success against a 90% release bar**.
- Designed provider interfaces and deterministic mocks so application tests run without a database, network or vendor keys. Integration tests requiring Postgres are separate.

Founder & Mobile Engineer · earlier products — Yoonibo LLC — Jul 2022 – Nov 2024

- Shipped **+Luv** (2022), a cross-platform couples app on the **App Store and Google Play** with WebSocket real-time chat, photo sharing, drawing notes and a subscription system, and **Jibbr** (2024), an OpenAI-powered conversation coach.

Software Developer — Fabspeed Motorsport — Jan 2022 – Dec 2022 · Philadelphia, PA

- Launched **wholesale.fabspeed.com**, built reusable React tools for bulk product updates, and wrote Python scrapers for Google and Yelp reviews.

SKILLS

Languages: Python, TypeScript, JavaScript, Swift, Kotlin, SQL

AI / LLM: RAG (BM25, embeddings, MMR), LLM evals and guardrails, LLM-as-judge, tool-calling agents, Claude, Gemini, OpenAI, local models (LM Studio, Ollama), Core ML, ONNX

Stack: Django, Fastify, PostgreSQL, Redis, SQLite, Docker, Heroku, GitHub Actions, React, Next.js, React Native / Expo

CERTIFICATIONS

TensorFlow Developer Certificate — image classification, NLP, and time-series forecasting